DOCUMENT RESUME

ED 321 751 IR 014 519

AUTHOR Schwier, Richard; Misanchuk, Earl R.
TITLE Analytical Tools and Research Issues for Interactive Media.

PUB DATE Jun 90
NOTE 18p.; Paper presented at the Annual Meeting of the Association for Media and Technology in Education in Canada (20th, St. John's, Newfoundland, June 9-12, 1990).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Cognitive Style; *Computer Assisted Instruction; *Courseware; Data Analysis; Foreign Countries; *Hypermedia; Individual Differences; *Instructional Design; *Interaction; Interactive Video; *Learner Controlled Instruction
IDENTIFIERS Audit Trails; *Path Analysis (Interactive Media)

ABSTRACT

Contrasting traditional linear media (in which all learners proceed through the instructional materials in an essentially fixed sequence) with interactive technologies and hypermedia (in which two learners may take different paths through the same instructional materials), this paper highlights the importance of recording and analyzing the instructional path taken by each learner, noting that such audit trail data can be used in the production of generalizable rules which explain the influences of social variables, individual differences, and cognitive style on the paths taken through the instruction. Descriptive and inferential approaches to analyzing interactive media audit trails are identified, and some of the issues and concerns raised by such investigation are discussed. The paper concludes with a consideration of research opportunities in interactive technologies of instruction that such analysis would promote. Samples of five methods of representing audit trails are appended. (9 references) (GL)

Analytical Tools and Research Issues for Interactive Media

Richard Schwier & Earl R. Misanchuk
The University of Saskatchewan
Saskatoon, SK. S7N 0W0

Presented at the annual meeting of the Association for Media and Technology in Education in Canada, St. John's, Newfoundland, June 9–12, 1990.

In traditional (linear) media, such as audio materials, video or film, or combinations of still pictures and audio, any learner going through the materials is offered essentially the same learning experience as any other learner going through the same materials (given reasonably similar levels of motivation). As a consequence, instructional designers have been able to assume a certain constancy to their instruction. Now, however, the proliferation of interactive technologies and hypermedia has given instructional designers the capability of producing materials which provide for a great deal of choice on the part of learners in determining their paths through the instruction. This means that two learners may be able to traverse the same instructional materials, perhaps even achieve the same objectives, yet not share a common experience. At minimum, the degree of commonality in the experience is reduced over that experienced with the more linear media. While obviously having a liberalizing (some would say humanizing) effect on instruction, the advent of interactive and hypermediated technologies brings with them a host of new questions for instructional designers, as well. Among these is the determination of the effects of taking different paths through instruction. To address such a question requires that there be some way of recording and analyzing the instructional path taken by each learner. We refer to that record of the path as the audit trail (after M. W. Petruk, personal communication, February 7, 1990).

It is a relatively simple matter to collect information about an individual's path, using contemporary computer-based

instruction authoring environments (e.g., Authorware Professional or Course of Action) or program development tools such as Macintosh's HyperCard. Learners' responses can be trapped and recorded at each decision point, or node. Of course, the descriptive data captured can provide very useful formative evaluation information. However, we speculate that there are additional purposes to which such information can be put, purposes that might produce generalizable rules about the construction of various paths through interactive and hypermediated instruction. In order to generate such rules, however, questions of this type must be asked and answered: Can we confidently state that one path is significantly different from another path? Can we combine groups of paths and compare them to other groups of paths—a necessary step if we intend to refine individual differences research in interactive media?

To illustrate the desirability of pursuing this type of research, consider that individuals, and indeed groups of individuals, can approach instruction differently, and this has traditionally been of theoretical interest. For a simple example, we might consider an issue like racism. Suppose an instructional developer is developing video materials for a Canadian (i.e., multicultural) audience. Might the race of an on-camera instructor influence choices made by a learner within the instructional sequence? What about other characteristics—socioeconomic status, gender, age, religion? Would any or all of these characteristics influence the ultimate path taken through the instruction?

As a second example, we could focus on any of a number of individual differences or cognitive styles; consider "locus of control." How might internalizers and externalizers differ in their approaches to highly organized interactive treatments? Would they react differently to, and take different paths through, very linear treatments and hypermedia treatments? One learner may select the shortest path available; another may select every available remedial segment in the same treatment. Resultant paths would be very different from one another, but, short of actually watching both individuals progress through the materials, how can these differences be expressed? As interest grows in the effectiveness of learner control of instruction (e.g., see Higginbotham-Wheat, 1990; López, 1990; Ross, Morrison, & O'Dell, 1990; Steinberg, 1977), and especially as it broadens into learner control of interactive and hypermediated instruction, these kinds of questions will command increasing interest.

In order to explore these types of questions, we need some analytical tools that don't appear to have been developed yet. As this area of investigation is still very new, prescriptions are not yet at hand. Some preliminary work on the subject has been done, however, and we attempt in this paper to represent in some meaningful ways the different paths that individuals may follow through interactive media treatments. We identify both descriptive and inferential approaches to analyzing interactive media audit trails, and identify some of the issues and concerns that such investigation reveals. We also attempt to identify research opportunities in interactive technologies of instruction that such analysis would promote.

The Audit Trail

An audit trail comprises all the responses generated by a learner going through interactive or hypermediated instruction. In general terms, audit trails contain words, phrases, sentences, and paragraphs that a learner types into a computer, as well as a record of the "multiple choice"-like responses made, either from the keyboard or via some other input device, such as a mouse or touch-sensitive screen. However, to keep this preliminary discussion relatively simple, we restrict our discussion here to the "multiple choice"-like responses. Hence, we conceive of the audit trail as a string of characters (numerals or letters) representing choices made by learners as they progress through choice points, or nodes, within the instruction. For example, suppose the first point at which a learner had to make a choice had three different paths the learner could follow, and let us further suppose that she chooses the second path. As she proceeds through the subsequent instruction, she encounters a second node, with two choices, and chooses the first of those. The first two characters of her audit trail would therefore be 21 (read "two-one", not "twenty-one"). If she then chooses the first of three paths at the next node, and the third of three at the following node, her complete audit trail up to this point would be 2113. A second learner might have chosen the third path at the first node, the first path at the second node, the first path at the third node, and the second path at the fourth node, to give an audit trail of 3112. In creating an audit trail, the computer would simply record the numerical value assigned to each option presented at a decision point, in a vector-like arrangement. As the number of decision points in the treatment increases, so will the length of the vector; as the complexity of the path taken by an individual increases (e.g., exploring optional paths as opposed to forging straight ahead), so will the length of the vector.

Some things immediately become obvious:

1. The meaning of any character in the audit trail is dependent for meaning upon the character preceding it. That is, the character '1', the second character in the first individual's audit trail, does not mean the same thing as the character '1' which is the second character in the second individual's audit trail, because the two individuals chose different paths at the first node. We will refer to this phenomenon as the dependency problem.

2. A problem is created by repeat visits to a node. In any particular interactive treatment, an individual may intersect the same node in the program more than once. Stated differently, the learner may traverse portions of a path repeatedly. Thus, if a simple frequency count is made of visits to a node, it is difficult to discriminate, for example, between the case in which two learners traverse the same path segment once each, and one in which a single learner traverses the same path segment twice. We will refer to this problem as the looping problem. The looping problem is context-dependent. Interactive treatments are typically designed to permit repeated visits to some portions of instruction but not others. A second implication of this phenomenon is that not all audit trails will be of the same length.

3. The problem of conceptual distance. This is a classical measurement problem which surfaces in the interpretation of data from interactive treatments. Given different treatments, selections at nodes may represent nominal, ordinal or even integer data points, and each type of data imposes restrictions on how data can be ana-

lyzed. For example, one menu in a program offers the viewer a choice of "river memories," "river bridges," "river travel," or "river science." In this case "river memories" is represented as 1 and "river science" as 4. Because the data are nominal, the numerical representation is misleading; the conceptual distance implied between the two choices is in fact no greater than the conceptual distance between any two other choices from the menu. The problem is not isolated to numerical data. With tree diagrams and other graphic approaches, conceptual distance is implied as branches diverge on a diagram. In actual fact, however, choosing "A" at the seventh level of the farthest branch on the right side of a diagram is not necessarily different conceptually from choosing "A" at the fourth level of the left branch. Nevertheless, a casual observer can be seduced into thinking that spatial or numerical distance in data indicate conceptual distance as well.

Descriptive Approaches to the Audit Trail

In our initial search for a meaningful way to represent the audit trail, we investigated and considered several formats. The list we generated, below, is not exhaustive, but merely a point of departure. Some of the representations appear to be more useful and durable than others; some we considered briefly and discarded for various reasons outlined below.

Raw Data Matrix

The most basic way of representing audit trail data is to simply record the responses of each learner (as a vector), one above the other (see Figure 1). Matrices like these have the advantage of being easily constructed and relatively easily interpreted for each individual. The interpretation, however, can only be relatively limited, and context-bound. For crude formative evaluation purposes, the data are useful. An instructional designer can see which choices are attracting individual learners, and speculate about design decisions. But there is a serious limitation with this type of approach: raw data buy definition, aren't summarized, and therefore conclusions based on the group are difficult to derive.

Nodal Frequencies and Proportions

Another approach we investigated was to present data associated with each decision point (node) in the treatment. Data can be presented in at least two forms, as raw data (Figure 2) or as proportions (Figure 3).

Raw nodal frequencies, like raw data matrices, are easy to create. They are perhaps easier to interpret, since the data are now summarized. Magnitudes of differences are obvious, and at any node, comparisons have high precision and are intuitively satisfying. At the same time, relationships across nodes, or among variables are difficult to interpret. For example, how should four choices of "A" at one node be compared with 103 choices of "A" at another node, if they occupy different locations in the treatment? Perhaps only eight individuals encountered the first node, whereas several

hundred encountered the second. The looping problems, described earlier in the paper, surface here to cause difficulties in interpretation. If an individual loops through a node several times, frequency data become distorted. Either the same choice is made several times, thereby inflating the frequencies at that node, or several different choices are made, thereby levelling the data at that decision point.

Nodal data can also be presented proportionally. Again, this type of format is easy to create and precision is retained at a high level. Proportions allow easier comparisons across nodes or among variables at different positions in the treatment. Of course some calculation is necessary, and the user must struggle with the question of what to use as a denominator. For example, is the denominator consistently the total number of learners encountering the treatment, or is the denominator the total number of learners who pass a particular decision point? Perhaps obviously, the denominator of any proportion will be determined by the comparisons the user chooses to make—yet another type of context dependence. As with raw nodal data, proportional data are also sensitive to looping problems. An individual looping several times through a particular node can inflate its proportion of the total. In addition, as one descends deeper into he data matrix, smaller raw numbers represent elevated proportions (see Figure 3). While proportional representation is useful for compressing large numbers, statistically bloating small numbers appears unnecessary and counterproductive.

When considering nodal data (either frequencies or proportions), individual differences are lost in the compression of data. Any design decisions based on these data are limited to conclusions about the group as a homogeneous entity, and we sacrifice any more subtle interpretations. Furthermore, data spread across several tables, each representing a single node (or, alternatively one large table showing multiple crossbreaks), are difficult to assimilate. Patterns that exist within them are difficult to detect.

Petit-Point Pattern

This early approach to the portrayal of the audit trail was suggested by some of John Tukey's work in representing non-parametric data in what he called a "stem and leaf" form (Hartwig & Dearing, 1979). It is a graphical approach, combining the intuitive appeal of a histogram with a character-based notational system symbolizing the choices. For example, in Figure 4, the X's represent the first choice, O's represent the second choice, and H's represent the third choice. A dot (•) is used to indicate that the learner proceeded past the node without making a choice, a situation made possible by an unfortunate bug in the program used to collect these sample data.

Although virtually any symbol could be used to represent individual choices, but we chose characters that seem to occupy the same amount of space, so that inter-character differences would not influence the overall appearance, and perhaps the interpretation, of the display. The resulting pattern of characters is in some ways reminiscent of a pattern for petit-point embroidery.

To set up a display of this kind, data must be progressively or sequentially sorted decision point by decision point. That is, subsequent columns of data in a matrix must be sorted within the categories formed by the sorted data comprising the first column.

Advantages of this method included the ease of generating the display by using a search and replace function on

a word processor to substitute characters, and a line sorting feature to assist in constructing the display. Although the display accurately shows the proportions of choices made at each node in a graphic and intuitive way, the dependency problem is very evidently in play: In the fifth column, for example, there are 11 distinct groups of O's; each group has a different meaning, depending on where it is located vertically. That is, although the O's all indicate that the second of the choices available was the one that was chosen, the first two O's and the third O represent choices on different content, due to the fact that different routes brought the learners to he node.

In one version of this kind of display, we also tried to use color to denote different choices, but found that it offered little advantage in interpretability.

This approac.ı was eventually set aside because it didn't seem to do a great deal to describe what was happening. It was deemed to be moderately useful as a formative evaluation tool, but didn't appear to have sufficient power to make it useful for research purposes.

Audit Trail Tree

The audit trail tree (Figure 5) was the next approach we attempted. It combined both the graphical representation and the numerical accuracy of the Petit-Point Pattern approach, but, in addition, presented the data in a more intuitively powerful way.

The audit trail tree is drawn so that the thickness of the line depicts the number of learners who chose the path represented. Of course, if large numbers of learners are involved, the line width could be scaled. Too, numbers (either frequencies or proportions) could be attached to each node to provide greater detail. The visual representation appeared to be useful and somewhat easier to interpret than the Petit-Point Pattern method. It was clear, it was graphical, it was intuitive, and it was grounded in reality. Comparisons were easy to make; flow could be read into the diagram as learners progressed from the beginning to the end of instruction. Although the drawing process is not difficult to do manually, it is somewhat tedious. and automation of the process on a graphic-interface microcomputer should be reasonably straightforward.

On the other hand, unless numbers were attached (as suggested above), the precision of the display was fairly low (i.e., it is difficult sometimes to tell the difference between 3 learners and 4, or between 11 and 13). The problem of conceptual distance remained, and perhaps was magnified by the ease with which other dimensions of the data were made manifest. And. of course, the problem of how to represent the loops remained.

Inferential Approaches to the Audit Trail

Inferential approaches would be used when comparisons are being made between groups of learners (e.g., differing on cognitive styles) or treatments (e.g., using instructors of the same gender as the learners and using instructors of differing genders).

One inferential approach to the audit trail we considered and discarded involves the use of multiple regression. This traditional approach to path analysis did not appear to be appropriate for answering the types of questions we are addressing here. It is a method of analyzing linear relationships among sets of variables, and assumes that a *causal order* among the variables is known and that the relationships among the variables are *causally closed* (Duncan, 1966). Even though we didn't give this approach much consideration, we mention it because of a possible confusion of terms: We are attempting to analyze paths through instruction, in a way that bears no relationship to path analysis, as the term is used in a statistical sense.

Furthermore, in an inferential approach to analyzing choices made in hypermediated and interactive instruction, one cannot make the assumption of normal distribution that underlies parametric statistics; hence a focus on the non-parametrics is essential.

A productive approach to analyzing a class of problems such as that under discussion would be to collect data on choices made by a large group of people, and regard that distribution as the *usual distribution* (in fact, it would be an expected distribution that is "normal" for the particular content and treatment being investigated, but since the term normal distribution has a technical connotation, we must make a distinction). Given this expected distribution, one could then subject certain individuals to treatments of varying kinds, and compare the audit trails of those subjects to the audit trails generated by the "usual" population, on a decision by decision basis (and keeping in mind the dependency problem). That is, the comparisons could only be made for single decision nodes at a time (which could be a limitation).

A statistic such as the χ^2 one-sample test, a test of goodness of fit (Siegel, 1956), would appear to be an appropriate tool for determining the statistical significance of observed deviations from "usuality". Another likely candidate would be the Kolmogorov-Smirnov test (Marascuilo & McSweeney, 1977; Siegel, 1956). Indeed, since Siegel states that "...the Kolmogorov-Smirnov test may in all cases be more powerful than...the χ^2 test" (1956, p. 51), it would seem to be the test of choice.

Issues and Challenges

There are issues and challenges arising in both the descriptive and inferential applications of audit trail analysis.

From the point of view of descriptive applications, an important issue underlying all the audit trail methods identified above is this: How does one deal with representing the fact that an individual learner may well traverse a node more than once? On one hand it would be possible to design tracking devices which would ignore subsequent identical path choices once an individual

has chosen the first time. But looping through the same path more than once may represent important data. How then can we discriminate between multiple encounters with a single individual across a path, and several single encounters with individuals? This problem remains unresolved, and deserves attention.

The highly contextual nature of all analyses continued to plague us as we searched for analytical tools. It seemed like every avenue we pursued was predicated on a thorough analysis of the treatment under study. The content not only determined to some extent which of the analytical tools were appropriate, but also limited our ability to generalize our conclusions at this stage of investigation. Certainly, we were humbled by the levels of complexity which seemed to settle like river sediment on what we initially thought was a simple notion.

The potential magnitude of an audit trail is also a challenge. Experience with collecting audit trail data in a variety of CAI and hypermedia environments (M. W. Petruk, personal communication, February 7, 1990) confirms the intuitive notion that extremely large datasets are possible. If one considers the potential number of choices available in an instructional sequence that involves any degree of learner choice, and factors in the possibility of looping through some or all nodes, the sheer size of the audit trails becomes daunting. Given that some learners will loop more than others, the audit trail data matrix will not be square; this in itself makes it awkward to work with.

In the inferential approach involving χ^2 or Kolmogorov-Smirnov, four distinct issues arise. The first has to do with the fact that any research must necessarily be conducted by comparing learner's responses at only one node at a time. If the number of nodes is large (as is liable to be the case in most instruction that incorporates learner control) then there exists the potential for an inflated α, similar to the situation where repeated *t*-tests are employed (Ryan, 1972). Other than recognizing the potential problem and selecting an appropriate α to compensate for it, no evident solution presents itself.

Another issue that arises in conjunction with the use of the χ^2 or Kolmogorov-Smirnov test: Which expected distribution should be used? Above, it was suggested that an empirically-derived expected distribution should be employed as the "usual". However, it could also be argued that a rectangular distribution, based on the number of paths available at each node, is the distribution of choice. More development of this argument, and identification of the pros and cons of using each distribution, are needed.

If it turns out that the empirically-derived distribution is the one of choice, then this question arises: How large does the population describing that distribution have to be in order to be adequate?

Finally, there is the opposite side of the coin: What should be the size of any sample of learners used in empirical research, given the knowledge that with larger sample sizes, χ^2 is know to become increasingly sensitive? That is, given a large sample size, the use of χ^2 may lead to incorrect decisions to reject H_0. How Kolmogorov-Smirnov behaves under these conditions is not clear.

Another consideration is that the nature of the content or subject matter may determine subsequent analyses, and we have not yet addressed this issue. In procedures we have dis-

cussed, the data do not represent the shapes of the instruction encountered. For example, they do not describe whether the learner is engaged in concept acquisition, problem solving or chaining at any particular point. This does not always matter, but where it is an important feature for analysis, the initial content analysis will need to take this into account. It is also possible that specific analytical procedures will need to be developed to address specific instructional types.

If we are to draw a single set of conclusions from our investigation of this issue to date, we suggest that the search for a single appropriate analytical tool for interactive media research is futile. We believe a thorough analysis of audit trail data requires a combination of techniques, both descriptive and inferential, to adequately address research questions. As our experience with existing tools grows, prescriptions concerning when to use combinations of approaches may be possible, but at this point, such prescriptions would be premature. Further, rather than viewing the context-dependence of each analysis as a barrier or limitation, we believe it is more productive to view content or task analysis as a prerequisite step to any further analysis.

References

Duncan, O.D. (1966). Path analysis: Sociological examples. *American Journal of Sociology*, 72, 1-16.

Hartwig, F., & Dearing, B. E. (1979). *Exploratory data analysis*. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-016. Beverly Hills, CA: Sage Publications.

Higginbotham-Wheat, N. (1990, February). *Learner control: When does it work?* Paper presented at the Annual Meeting of the Association of Educational Communications and Technology, Anaheim, CA.

López, C. L. (1990, February). *Personalizing math word problems*. Paper presented at the Annual Meeting of the Association of Educational Communications and Technology, Anaheim, CA.

Marascuilo, L. A., & McSweeney, M. (1977). *Nonparametric and distribution-free methods for the social sciences*. Monterey, CA: Brooks/Cole Publishing.

Ross, S., Morrison, G., & O'Dell, J. (1990, February). *Uses and effects of learner control of context and instructional support in computer-based instruction*. Paper presented at the Annual Meeting of the Association of Educational Communications and Technology, Anaheim, CA.

Ryan, T. A. (1972). Multiple comparisons in psychological research. In R. E. Kirk (Ed.), *Statistical issues: A reader for the behavioral sciences* (pp. 291–306). Monterey, CA: Brooks/Cole Publishing.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.

Steinberg, E. R. (1977). Review of student control in computer-assisted instruction. *Journal of Computer-Based Instruction*, 3(3), 84–90.

Subject 1	2	2	2	2	1	1	1	1
Subject 2	2	2	2	2	1	1	1	1
Subject 3	2	2	2	2	1	1	2	1
Subject 4	2	2	2	2	1	1	2	1
Subject 5	2	1	2	2	1	1	1	1
Subject 6	1	3	2	2	2	1	1	1
Subject 7	2	2	2	2	2	1	1	1
Subject 8	2	1	2	2	2	1	1	1
Subject 9	2	3	2	2	1	1	2	2
Subject 10	2	1	2	2	1	1	2	2
Subject 11	2	3	2	2	1	1		
Subject 12	2	3	4	4	1	1		
Subject 13	2	3	4	4	1	1		
Subject 14	2	3	4	4	1	1		
Subject 15	2	2	2	2	1	1	1	
Subject 16	2	2	4	2	1	1		
Subject 17	1	1	2	2	1	1		
Subject 18	2	1	2	2	1	1		
Subject 19	2	1	2	2	1	1		
Subject 20	2	1	2	4	1	1		
Subject 21	1	3	4	1	2	1	2	
Subject 22	1	3	4	1	2	1	2	
Subject 23	1	3	4	4	2	[illegible]		
Subject 24	2	3	4	2	2	1	2	
Subject 25	2	2	2	2	2	1	1	
Subject 26	2	2	4	2	2	1	2	
Subject 27	2	1	4	2	2	1	2	
Subject 28	1	3	2	2	1	2	1	1
Subject 29	2	3	2	2	1	2	1	1
Subject 30	2	2	2	2	1	2	1	1
Subject 31	2	2	2	2	1	2	1	1
Subject 32	2	2	2	2	1	2	1	1
Subject 33	2	1	2	2	1	2	1	1
Subject 34	2	3	2	2	2	2	1	1
Subject 35	2	3	2	2	2	2	1	1
Subject 36	2	3	2	2	2	2	1	1
Subject 37	2	2	2	2	2	2	1	1
Subject 38	2	2	2	2	2	2	1	1
Subject 39	2	2	2	2	2	2	2	1
Subject 40	2	1	2	2	2	2	1	1
Subject 41	2	1	2	2	2	2	2	1
Subject 42	1	3	2	2	1	2	2	2
Subject 43	2	3	2	2	1	2	1	2
Subject 44	2	3	2	2	1	2	1	2
Subject 45	2	3	2	2	1	2	2	2
Subject 46	2	3	2	2	1	2	2	2
Subject 47	2	1	2	2	1	2	2	2
Subject 48	2	3	2	2	2	2	1	2
Subject 49	2	3	2	2	2	2	2	2
Subject 50	2	3	2	2	2	2	2	2
Subject 51	2	2	2	2	2	2	1	2
Subject 52	2	2	2	2	2	2	1	2
Subject 53	2	2	2	2	2	2	1	2
Subject 54	2	2	2	2	2	2	2	2
Subject 55	1	1	2	2	2	2	2	2
Subject 56	2	1	2	2	2	2	1	2
Subject 57	2	1	2	2	2	2	2	2
Subject 58	2	3	2	2	1	2		
Subject 59	2	3	4	2	1	2		
Subject 60	2	3	4	4	1	2		
Subject 61	2	3	4	4	1	2		
Subject 62	1	2	4	2	1	2		
Subject 63	2	2	2	2	1	2		
Subject 64	2	2	2	2	1	2		
Subject 65	2	1	2	2	1	2		
Subject 66	2	1	2	2	1	2		
Subject 67	2	1	2	2	1	2		
Subject 68	1	3	4	1	2	2	2	
Subject 69	1	3	4	2	2	2	2	
Subject 70	1	3	4	4	2	2		
Subject 71	1	3	4	4	2	2		
Subject 72	1	3	4	4	2	2		
Subject 73	2	3	2	2	2	2		
Subject 74	2	3	2	2	2	2		
Subject 75	2	3	2	2	2	2		
Subject 76	2	3	4	4	2	2		
Subject 77	2	3	4	4	2	2		
Subject 78	2	3	4	4	2	2		
Subject 79	2	3	4	4	2	2		
Subject 80	2	3	4	4	2	2		
Subject 81	2	3	4	4	2	2		
Subject 82	2	3	4	4	2	2		
Subject 83	2	3	4	4	2	2		
Subject 84	2	3	4	4	2	2		
Subject 85	2	3	4	4	2	2		
Subject 86	2	3	4	4	2	2		
Subject 87	2	3	4	4	2	2		
Subject 88	2	3	4	4	2	2		
Subject 89	2	3	4	4	2	2		
Subject 90	2	2	4	1	2	2	2	
Subject 91	1	1	2	2	2	2		
Subject 92	2	1	2	2	2	2		
Subject 93	2	1	2	2	2	2		
Subject 94	2	1	2	2	2	2		
Subject 95	2	3	2	2	1	3	2	1
Subject 96	2	1	2	2	1	3	1	1
Subject 97	2	1	2	2	1	3	2	1
Subject 98	2	1	2	2	2	3	1	2
Subject 99	2	1	2	2	2	3	2	2
Subject 100	2	2	2	2	2	4		
Subject 101	1	3	4	4	1			
Subject 102	1	2	4	1	1			
Subject 103	1	2	4	1	1			
Subject 104	1	2	4	2	1			
Subject 105	2	3	4	4	2			
Subject 106	1	2	4	2	2			
Subject 107	1	4	4	4	4			
Subject 108	1	4	4	4	4			
Subject 109	1	4	4	4	4			
Subject 110	1	4	4	4	4			
Subject 111	1	4	4	4	4			
Subject 112	2	4	4	4	4			
Subject 113	2	4	4	4	4			
Subject 114	2	4	4	4	4			
Subject 115	2	4	4	4	4			
Subject 116	2	4	4	4	4			
Subject 117	2	3	4	4	4			
Subject 118	1	2	2	2	4			
Subject 119	1	2	4	1	4			
Subject 120	1	2	4	2	4			
Subject 121	2	2	4	2	4			
Subject 122	2	2	4	2	4			
Subject 123	1	1	2	2	4			
Subject 124	1	1	2	2	4			
Subject 125	1	1	2	2	4			
Subject 126	1	1	2	4	4			
Subject 127	2	1	2	2	4			
Subject 128	2	1	2	2	4			
Subject 129	2	1	2	4	4			
Subject 130	2	1	2	4	4			
Subject 131	2	1	2	4				

Figure 1. Sample raw data matrix.

<table>
<tr><td rowspan="2">Node 1</td><td>Resp.</td><td colspan="24">1</td></tr>
<tr><td>Freq.</td><td colspan="24">32</td></tr>
<tr><td rowspan="2">Node</td><td>Resp.</td><td colspan="6">1</td><td colspan="6">2</td><td colspan="6">3</td><td colspan="6">0</td></tr>
<tr><td>Freq.</td><td colspan="6">7</td><td colspan="6">8</td><td colspan="6">12</td><td colspan="6">5</td></tr>
<tr><td rowspan="2">Node 3</td><td>Resp.</td><td colspan="3">2</td><td colspan="3">0</td><td colspan="3">2</td><td colspan="3">0</td><td colspan="3">2</td><td colspan="3">0</td><td colspan="3">2</td><td colspan="3">0</td></tr>
<tr><td>Freq.</td><td colspan="3">7</td><td colspan="3">0</td><td colspan="3">1</td><td colspan="3">7</td><td colspan="3">3</td><td colspan="3">8</td><td colspan="3">0</td><td colspan="3">5</td></tr>
<tr><td rowspan="2">Node 4</td><td>Resp.</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td><td>0</td></tr>
<tr><td>Freq.</td><td>0</td><td>8</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>3</td><td>4</td><td>0</td><td>0</td><td>3</td><td>0</td><td>3</td><td>1</td><td>5</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>5</td></tr>
<tr><td rowspan="2">Node 5</td><td>Resp</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Freq.</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

<table>
<tr><td rowspan="2">Node 1</td><td>Resp.</td><td colspan="24">2</td></tr>
<tr><td>Freq.</td><td colspan="24">99</td></tr>
<tr><td rowspan="2">Node</td><td>Resp.</td><td colspan="6">1</td><td colspan="6">2</td><td colspan="6">3</td><td colspan="6">0</td></tr>
<tr><td>Freq.</td><td colspan="6">28</td><td colspan="6">25</td><td colspan="6">41</td><td colspan="6">5</td></tr>
<tr><td rowspan="2">Node 3</td><td>Resp.</td><td colspan="3">2</td><td colspan="3">0</td><td colspan="3">2</td><td colspan="3">0</td><td colspan="3">2</td><td colspan="3">0</td><td colspan="3">2</td><td colspan="3">0</td></tr>
<tr><td>Freq.</td><td colspan="3">27</td><td colspan="3">1</td><td colspan="3">20</td><td colspan="3">5</td><td colspan="3">18</td><td colspan="3">23</td><td colspan="3">0</td><td colspan="3">5</td></tr>
<tr><td rowspan="2">Node 4</td><td>Resp.</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td><td>0</td></tr>
<tr><td>Freq.</td><td>0</td><td>23</td><td>4</td><td>0</td><td>1</td><td>0</td><td>0</td><td>20</td><td>0</td><td>1</td><td>4</td><td>0</td><td>0</td><td>18</td><td>0</td><td>0</td><td>2</td><td>21</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>5</td></tr>
<tr><td rowspan="2">Node 5</td><td>Resp</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Freq.</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Figure 2. Tabular representation of raw frequencies of paths chosen in sample data. Although the sample data actually have eight decision nodes, only four are presented here, for simplicity. Note how quickly the size and complexity of the table grows as the number of decision nodes and/or the number of choices available at each node increases.

Node 1	Resp.	1																							
	Prop.	24																							
Node 2	Resp.	1						2						3						0					
	Prop.	22						25						38						15					
Node 3	Resp.	2			0			2			0			2			0			2			0		
	Prop.	100			0			13			87			25			75			0			100		
Node 4	Resp.	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0
	Prop.	0	86	14	0	0	0	0	100	0	42	58	0	0	100	0	33	[illegible]	56	0	0	0	0	0	100
Node 5	Resp.	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I
	Prop.	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I

Node 1	Resp.	2																							
	Prop.	76																							
Node 2	Resp.	1						2						3						0					
	Prop.	28						25						42						5					
Node 3	Resp.	2			0			2			0			2			0			2			0		
	Prop.	96			4			80			20			44			56			0			100		
Node 4	Resp.	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0
	Prop.	0	85	15	0	100	0	0	100	0	20	80	0	0	100	0	0	9	81	0	0	0	0	0	100
Node 5	Resp.	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I
	Prop.	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I	I I I

Figure 3. Tabular representation of proportions of paths chosen in sample data for first four decision nodes. Note how, as one progresses through more decision nodes, smaller frequencies of responses produce deceptively larger proportions.

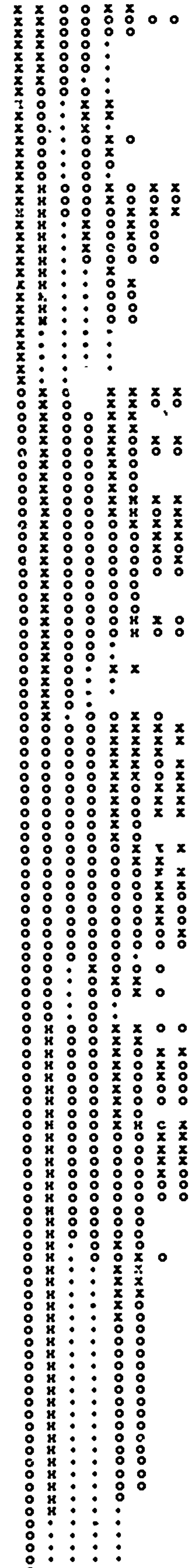

Figure 4. Petit-Point Pattern method of representing an audit trail. An X represents the choice of the first possible path from a node, an O represents the second, and an H represents the third. A • indicates no response.

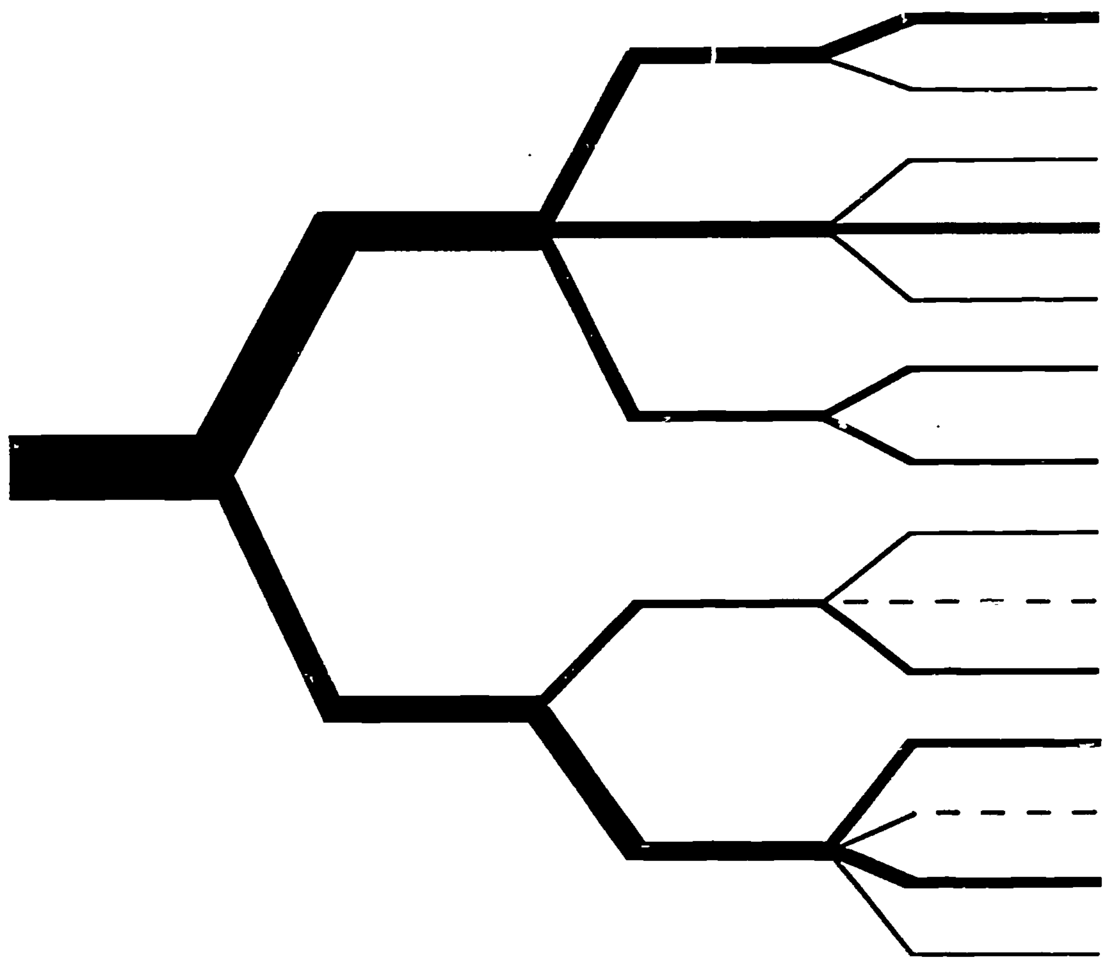

Figure 5. Audit Trail Tree of example data. The width of the line represents the number of learners taking any given path. A dashed line indicates no learners took the path.